

## Development of regional regression relationships with censored data

Charles N. Kroll

Environmental Resources and Forest Engineering, College of Environmental Science and Forestry  
State University of New York, Syracuse

Jery R. Stedinger

School of Civil and Environmental Engineering, Cornell University, Ithaca, New York

**Abstract.** When no discharge record is available for a site, a regional regression relationship can be used to estimate low-flow quantiles. Problems arise in the derivation of such models when some at-site quantile estimates are reported as zero. One concern is that quantile estimates reported as zero may be in the range from zero to the measurement threshold. A second concern is that a logarithmic transformation cannot be used with zero quantile estimates, so traditional log linear least squares estimators cannot be computed. This paper uses visual examples and Monte Carlo simulation to compare the performance of techniques for estimating the parameters of a regional regression model when some at-site quantile estimates are zero. Ordinary least squares (OLS) techniques employed in practice include adding a small constant to all at-site quantile estimates (denoted OLSC), or neglecting all observation reported as zero (denoted OLSD). OLSC and OLSD performed poorly compared to the use of a Tobit model, which is a maximum likelihood estimator (MLE) procedure that represents the below threshold estimates as a range from zero to the threshold level. For a small amount of censoring, the OLSD method can be acceptable. A weighted Tobit model that accounts for the heteroscedasticity of the residuals in the regression model provided relatively little gain over the ordinary Tobit model.

### 1. Introduction

Estimates of low-streamflow statistics are needed for planning hydropower, irrigation and water supply systems, cooling-plant facilities, waste-load allocations into streams, and recreational uses. When river discharge records are available, low-streamflow statistics can be obtained by either a frequency analysis [Riggs, 1968, 1972] or cross-correlation techniques using a longer record site [Hirsh, 1979, 1982; Vogel and Stedinger, 1985; Stedinger and Thomas, 1985], depending on the length of record available at the site of interest. When no discharge record exists for a site, a regional regression model can be used to estimate low-streamflow statistics.

Relationships between low-flow statistics and geomorphic, geologic, climatic, and topographic parameters have been developed for many regions [Thomas and Benson, 1970; Thomas and Cervione, 1970; Riggs, 1972; Parker, 1977; Bingham, 1986; Vogel and Kroll, 1992; Dingman and Lawlor, 1995]. These models often have the form

$$Q_{d,t} = e^{\alpha} X_1^{\beta} X_2^{\gamma} \cdots \quad (1)$$

where  $Q_{d,t}$  is the day ( $d$ ) year ( $t$ ) low-flow statistic,  $X_i$  are drainage basin characteristics, and  $\alpha$ ,  $\beta$ , and  $\gamma$  are model parameters. Taking the logarithm of both sides of (1) yields a linear regression problem. The parameters in this model can be estimated using at-site quantile estimates from gauged river sites that provide estimators of  $Q_{d,t}$ . Once a model has been

developed for a region, low-flow estimates at ungauged sites can be obtained as a function of drainage basin characteristics.

At some gauged sites, at-site quantile estimates are reported as zero. Such low-flow estimates occur at stations where the river discharge is sometimes reported as zero. At times the flow in the river may actually be zero, while in other instances flows may be nonzero but too small to be recorded by the available measuring instrumentation and consequently recorded as zero. An analysis of (1) is inconsistent with zero quantile estimates because zero quantile estimates may represent a range from zero to the measurement threshold and not a single value. In addition, zero quantile estimates are incompatible with the logarithmic transformation commonly applied to solve for the parameters in (1). Recognizing that quantiles below some threshold value cannot be distinguished from zero also eliminates many problems associated with the estimation of a lower bound of some fitted frequency distribution [Lawal and Watt, 1996].

The situation where all data below a fixed value is censored is called type I censoring, as opposed to type II censoring, where a fixed number of data points are always censored [David, 1981]. Censored low-flow regional regression models should be an example of type I censoring because the measurement threshold at a river site is fixed, and all flows less than the measurement threshold are reported as zero. Two simple methods have been proposed to deal with type I censoring in low-flow regional regression models. The first is illustrated by Hammett [1984] who added a constant to all low-flow estimates in real space before performing ordinary least squares (OLS) regression on the translated quantiles in logarithmic space.

Copyright 1999 by the American Geophysical Union.

Paper number 98WR02743.  
0043-1397/99/98WR-02743\$09.00

The second is illustrated by *Arihood and Glatfelter* [1991] who neglected all observations below 0.05 cfs (1 cfs is  $2.8317 \times 10^{-2}$  m<sup>3</sup>/s) and performed OLS regression on the remaining data set. *Ludwig and Tasker* [1993] employed a more sophisticated approach by using logistic regression to model the probability of zero flows at a site. They then combined this information with a regional regression model for the mean and variance of the flows which was developed using only sites with nonzero at-site quantile estimates.

Nonlinear least squares fit in real space could be used to estimate the parameters in (1) [McCuen *et al.*, 1990]. Such a procedure has a number of problems. Unweighted least squares in real space misrepresents the error structure of the problem which one should capture to obtain statistically efficient estimators. This is an important issue. In many regions the at-site quantile estimates,  $Y_i$ , as well as some drainage basin characteristics (such as drainage area),  $X_i$ , vary by several orders of magnitude. If one gives the residuals from each site equal weight, information at sites with smaller quantile estimates would be lost. A logarithmic transformation appears to provide a good description of the residual error in hydrologic regional regression models and has been applied in practice for low flows [Thomas and Benson, 1970; Vogel and Kroll, 1992] and flood flows [Jennings *et al.*, 1994]. A similar result might be obtained by weighted least squares in real space. However, that approach has other problems, particularly with censored data. To obtain unbiased parameter estimators, one needs to account for the asymmetry of the truncated error distribution, which nonlinear least squares in real space does not. In addition, zero quantile estimates represent a range from zero to the measurement threshold. If reported zeros are treated as zeros, the resulting model may incorrectly describe the likelihood of low streamflows in small basins. Because of these problems, a real-space nonlinear least squares model is not included in this experiment.

*Tobin* [1958] considered representing the censored population as a mixed distribution, where the censored observations represent an interval of the sample space less than the detection limit. A maximum likelihood technique is used to estimate the model parameters. This model has been called a censored regression model in the statistics literature [Judge *et al.*, 1985] and a Tobit model in the econometrics literature [Amemiya, 1985]. Recently, *Liu et al.* [1997] applied a Tobit model to environmental quality data, and *Lu et al.* [1998] considered a bivariate censored model for low-streamflow prediction.

Few studies have compared the performance of different regional regression techniques for handling censored hydrologic data. *Kroll and Stedinger* [1994] compared a number of different regression models for censored data using Monte Carlo simulation. Their work was an extension of *Liu and Stedinger's* [1991] analysis which appears to contain errors in the parameter estimation procedure for the Tobit model. *Kroll and Stedinger* used the combined mean square error (mse) of all quantile estimators above and below the censoring threshold as a performance criterion to compare estimation techniques. They found the quantile estimator for the Tobit method to have a smaller mse than the quantile estimators based on regression after adding a constant or ignoring the censored observations. In general, quantile estimators based on a model which ignores the censored observations had a smaller mse than quantile estimators based on a model which adds a constant to all at-site quantile estimates, especially when the model error variance was small.

The underlying two-parameter regression model in *Kroll and Stedinger's* [1994] and *Liu and Stedinger's* [1991] Monte Carlo experiment had a relatively shallow slope ( $\beta \cong 0.5$  in (1)) compared to many low-flow regional regression models [Thomas and Benson, 1970]. Of interest here is whether the same conclusions would be drawn when estimators of only the above threshold quantile estimators are compared for models with more realistic slopes. Quantile estimators below the censoring threshold would generally be described as equal to zero, and how well the estimation techniques fit these values is likely to be of little practical importance. The gains from implementing a weighted Tobit model which accounts for the heteroscedastic nature (nonconstant variance) of the residuals in the regression model is also investigated.

## 2. Estimation Techniques

A simple two-parameter regression model is considered in these experiments. The dependent variable is the 7-day, 10-year low-flow,  $Q_{7,10}$ , the most widely used low-flow index in the United States [Riggs *et al.*, 1980]. Drainage area, a commonly used descriptor of low-flow quantiles in regional regression models, is used as the independent variable in these experiments. Four estimation procedures are analyzed. The first (OLSC) adds a constant to all at-site quantile estimators before performing ordinary least squares (OLS) regression. The second (OLSD) ignores the censored observations and performs OLS regression only with the uncensored observations. The third is a Tobit model which is a maximum likelihood estimation (MLE) technique that represents the censored observations as a range from zero to the measurement threshold. The fourth is a weighted Tobit model which accounts for the heteroscedasticity (nonconstant variance) of the residuals in regional regression analyses.

### 2.1. OLSC Method

The two-parameter regional regression model in logarithmic form is

$$\ln(Q_i) = \alpha + \beta \ln(A_i) + \varepsilon_i \quad (2)$$

where  $Q_i$  is an at-site estimator of the  $Q_{7,10}$  at site  $i$ ,  $A_i$  is the drainage area, and  $\varepsilon_i$  is the residual error.  $Q_i$  is obtained using the available discharge record at site  $i$ . To employ this model, all  $Q_i$  estimates must be greater than zero, which is not always the case [Hammett, 1984; Arihood and Glatfelter, 1991; Ludwig and Tasker, 1993]. To avoid taking the logarithm of zero, one could add a constant,  $c$ , to all at-site quantile estimates before performing the regression analysis [Hammett, 1984]. The dependent variable in the regression model becomes

$$Y_i = \ln(Q_i + c) \quad (3)$$

so that the regression equation is

$$Y_i = \alpha + \beta \ln(A_i) + \varepsilon_i \quad (4)$$

If the error terms are assumed to be independent and identically distributed (iid) with mean zero and constant variance, the parameters in (4) can be estimated efficiently by OLS regression [Johnston, 1972]. Once the regression parameters have been estimated as  $\hat{\alpha}$  and  $\hat{\beta}$  these estimates can be used to obtain a quantile estimate at site  $i$  as

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta} \ln(A_i) \quad (5)$$

The constant added to all at-site estimates should be subtracted from the quantile estimates produced by this model to obtain quantile estimates for a particular site:

$$\hat{Q}_i = \exp(\hat{Y}_i) - c \quad (6)$$

## 2.2. OLS Method

Another approach would be to neglect all sites with zero quantile estimates in the regression analysis [Arihood and Glatfelter, 1991]. OLS regression is then performed on the remaining uncensored data. This model can be written

$$\ln(Q_i) = \alpha + \beta \ln(A_i) + \varepsilon_i \text{ if } Q_i > 0 \quad (7)$$

This procedure reduces the number of data points in the analysis and thus suffers from some loss of information and possible bias [see Amemiya, 1985].

## 2.3. TOBIT Method

The Tobit model can be written

$$\begin{aligned} Y_i = \ln(Q_i) &= \alpha + \beta \ln(A_i) + \varepsilon_i & \text{if } Q_i > T \\ Y_i &= \ln(Q_0) & \text{otherwise} \end{aligned} \quad (8)$$

where  $\varepsilon_i$  are assumed to be independent, normally distributed residual errors with mean zero and constant variance,  $\sigma_\varepsilon^2$ ,  $T$  is the censoring threshold value, and  $Q_0$  is some nominal value  $\leq T$ . For low-flow quantiles,  $Q_0$  is often considered equal to zero. The likelihood function for this model is

$$L = \prod_{i=1}^M F_i(T) \prod_{i=M+1}^N f_i(Y_i) \quad (9)$$

where  $M$  is the number of sites with at-site quantile estimates equal to zero and  $i = M + 1, \dots, N$  corresponds to the sites with uncensored observations.  $F_i(T)$  is the probability  $Y_i$  is below the threshold value and  $f_i(Y_i)$  is the density function of  $Y_i$  when  $Q_i$  is greater than  $T$ :

$$F_i(T) = \int_{-\infty}^{(T - \alpha - \beta \ln(A_i)) / \sigma_\varepsilon} \left[ \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \right] dx \quad (10)$$

$$f_i(Y_i) = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\left\{-\frac{1}{2} \left[ \frac{Y_i - \alpha - \beta \ln(A_i)}{\sigma_\varepsilon} \right]^2\right\} \quad (11)$$

The assumed value,  $Q_0$  of  $Y_i$  when  $Q_i$  is less than  $T$  has no effect upon the likelihood function.

If the error terms in the Tobit model are assumed to be homoscedastic and independent, the model parameters may be efficiently estimated using maximum likelihood estimation techniques [Amemiya, 1985]. By taking the logarithm of (9) and setting the partial derivative with respect to the model parameters,  $\alpha$ ,  $\beta$ , and  $\sigma_\varepsilon$ , to zero, one can attempt to solve for the parameters in the model. Amemiya [1973] proved that the log-likelihood function of the Tobit model is not globally concave when solved with respect to the parameters in this form; thus maximization techniques may not converge on a global maximum. A suggested transformation is to take the partial derivative of the log-likelihood function with respect to

$$\alpha^* = \frac{\alpha}{\sigma_\varepsilon} \quad \beta^* = \frac{\beta}{\sigma_\varepsilon} \quad h = \frac{1}{\sigma_\varepsilon} \quad (12)$$

Olsen [1978] proved that the log-likelihood function of the Tobit model is globally concave after this transformation. Olsen [1978] provides the first and second partial derivative with respect to the expression in (12), and thus they are not included here. Since the log-likelihood function is globally concave, optimization techniques such as Newton's method [Chapra and Canale, 1985] can be employed to solve for the model parameters.

## 2.4. Weighted Tobit Method

The likelihood function used with the TOBIT method represents the residual error terms as independent and identically distributed normal random variables. This condition is often violated due to variations in the magnitude of the sampling error associated with the at-site quantile estimators and cross-correlation among the flows from which the at-site quantile estimators are computed. The differences in sample error variance are primarily due to differences in record lengths. Tasker [1980], Stedinger and Tasker [1985, 1986a, b], and Tasker and Stedinger [1989] developed regional regression techniques to handle such situations in the absence of censoring.

Our weighted Tobit method addresses the varying sample error variance in the at-site quantile estimators. The likelihood function for the weighted Tobit method, WTOBIT, is the same as in (9) if  $\sigma_\varepsilon^2$ , the residual error variance, is replaced by  $\gamma^2 + \Sigma_{ii}$  in the equations for the distribution and density function (equations (10) and (11)). Here  $\gamma^2$  is the underlying model error variance corresponding to  $\sigma_\varepsilon^2$  in the ordinary Tobit model, and  $\Sigma_{ii}$  is the sampling error of the at-site quantile estimators. Parameters  $\gamma$ ,  $\alpha$ , and  $\beta$  are estimated using the maximum likelihood approach. Similar to the TOBIT method, a transformation of the model parameters of the log-likelihood function was employed:

$$\alpha^* = \frac{\alpha}{\gamma} \quad \beta^* = \frac{\beta}{\gamma} \quad h = \frac{1}{\gamma} \quad (13)$$

Construction of  $\Sigma_{ii}$  to be employed with the WTOBIT estimator is discussed in section 3. In some applications, data sets yielding  $\gamma = 0$  might be a concern; however, because the model error variance of low-flow regional regression models tends to be high [Thomas and Benson, 1970], this situation is unlikely.

## 3. Experimental Design

The estimation methods discussed in section 2 are compared in several Monte Carlo experiments. This section describes (1) the regional model used to generate data in the Monte Carlo simulation, (2) at-site quantile estimation when some annual minimum flows are reported as zero, (3) the censoring of at-site quantile estimates, (4) quantile estimation using the fitted regression models, (5) construction of the weighting matrix  $\Sigma$  for the WTOBIT method, and (6) the performance measures used to compare the regression methods.

### 3.1. Underlying Model

To examine the performance of the quantile estimators associated with the different estimation methods using a Monte Carlo experiment, it is necessary to specify a true regional model. Annual minimum 7-day low flows at a site are assumed to be well described by a lognormal distribution. Vogel and Kroll [1989] and Dingman and Lawlor [1995] showed that this is a reasonable model of annual minimum flows in the north-eastern United States. Tasker [1987] and Durrans and Tomic

[1996] both recommended a log-Pearson III distribution to fit low-streamflow series in Virginia and Alabama, respectively. The lognormal distribution is a special case of the log-Pearson III distribution.

Each experiment included 25 sites with the drainage areas evenly spaced in logarithmic units from 10 to 300 square miles. At each site, the mean,  $\mu_i$ , and standard deviation,  $\sigma_i$ , of the annual minimum 7-day low flows in logarithmic space were calculated as a function of drainage area using the models

$$\mu_i = \alpha_\mu + \beta_\mu \ln(A_i) + \varepsilon_i \quad (14)$$

$$\sigma_i = \alpha_\sigma + \beta_\sigma \ln(A_i) \quad (15)$$

where  $\varepsilon_i$  are independent normally distributed random errors with mean zero and constant variance,  $\sigma_\varepsilon^2$ . It is assumed that the variance of the error in (15) is relatively small and thus can be neglected. *Kroll and Stedinger* [1998] and *Stedinger and Tasker* [1985] show that a small error term in (15) results in very little heteroscedasticity and thus has a negligible impact on the analysis.

Assuming the annual minimum flows are lognormally distributed, the underlying quantile model in each experiment is

$$Y_i = \alpha + \beta \ln(A_i) + \varepsilon_i \quad (16)$$

where  $Y_i$  is a quantile with nonexceedence probability  $p$ ,  $\alpha = \alpha_\mu + z_p \alpha_\sigma$ ,  $\beta = \beta_\mu + z_p \beta_\sigma$ , and  $z_p$  is the frequency factor associated with the 100 $p$  percentile of a standard normal distribution. The variance of the residual errors,  $\sigma_\varepsilon^2$ , corresponds to the model error variance in (16). Parameter  $\sigma_\varepsilon$  was assigned values of 0.3, 0.5, and 0.7 to examine how variations in the model error variance effect the performance of the quantile estimators. The values of  $\sigma_\varepsilon$  correspond to standard errors of estimate in real space ranging from 30 to 80%. These values represent small to large model error variances for realistic regional low-flow regression models [*Thomas and Benson*, 1970; *Vogel and Kroll*, 1990, 1992].

Models with two different slopes are considered in this experiment. In the shallow slope model the coefficients  $\beta_\mu$ ,  $\beta_\sigma$ , and  $\alpha_\sigma$  were set to 0.35,  $-0.12$ , and  $1.1$ , respectively; this is consistent with the model considered by *Liu and Stedinger* [1991]. These values correspond to  $\beta = 0.504$  in (16), which means that  $Q_{7,10}$  increases linearly with  $A^{0.504}$ . The parameter  $\alpha_\mu$  was set to four different values to vary the fraction of data which falls below the threshold. The selected values,  $\alpha_\mu = -2.2, -2.5, -2.8$ , and  $-3.1$ , correspond to censoring on average of 3, 7, 11, and 15 of the 25 observations.

The coefficient  $\beta = 0.504$  appears to be a small slope for a two-parameter regional regression model. Using the Boussinesq equation, *Vogel and Kroll* [1992] developed a physical model for low flows based on flow into a fully penetrating stream channel from an unconfined rectangular aquifer placed on a horizontal impermeable layer. For this multivariate physical model, the coefficient for drainage area  $\beta$  equals one. However, with a two-parameter model employing a single explanatory variable as in (16), the explanatory variable becomes a surrogate for other variables (such as average basin slope, hydrogeologic properties, precipitation, etc.), and thus the exponent may take on other values. *Thomas and Benson* [1970] developed regional regression models for the 7-day, 2-year and 7-day, 20-year low flows and generally found the coefficient on drainage area to be in the range from 1 to 2 for models with drainage area as the only explanatory variable.

In the medium slope model, the coefficients  $\beta_\mu$ ,  $\beta_\sigma$ , and  $\alpha_\sigma$  were set to 1.03,  $-0.08$ , and  $0.37$ , respectively. These values are consistent with those found by *Thomas and Benson* [1970] for low-flow regional regression models and correspond to  $\beta = 1.132$  in (16). To vary the amount of censoring, the parameter  $\alpha_\mu$  was assigned four different values,  $-4.9, -5.4, -6.1$ , and  $-6.7$ , which again correspond to censoring on average of 3, 7, 11, and 15 of the 25 observations.

### 3.2. Data Generation and At-Site Quantile Estimation

To implement the censored regression methods discussed in section 2, one must first estimate quantiles at each site using available discharge records. A record of random log annual minimum 7-day flows of length  $n_i$  was generated for each site using a normal distribution with mean and variance given by (14) and (15), respectively. The record length at 15 sites was 10 years, and the record length at 10 sites was 50 years. The measurement threshold could vary across sites, as it may in practice. Here it was set to 0.1 cfs at all sites. If the measurement threshold varied across sites,  $T_i$  would replace  $T$  in (9) and (10) for the Tobit method, where  $T_i$  is the measurement threshold at site  $i$ . Otherwise, implementation of this method would not change.

At each site, a random number,  $n_c$ , of annual minimum flows will be below the measurement threshold value. These values are censored; in practice they would be reported as zero. Each annual minimum flow above the measurement threshold was assigned a plotting position calculated as

$$p_j = \frac{n_c}{n_i} + \left( \frac{n_i - n_c}{n_i} \right) \left[ \frac{(j - 3)/(8 - n_c)}{n_i - n_c + 1/4} \right] \quad j \geq n_c + 1 \quad (17)$$

where  $j$  is the rank of the  $j$ th flow. For the data above the measurement threshold the logarithm of the annual minimum flows,  $y_i$ , are regressed against the corresponding "normal scores" using the model

$$y_j = \hat{\mu}_i + \hat{\sigma}_i \Phi^{-1}(p_j) + \varepsilon_j \quad (18)$$

where  $\hat{\mu}_i$  and  $\hat{\sigma}_i$  are the resulting estimators of the mean and standard deviation of the logarithm of the annual minimum 7-day flows obtained by ordinary least square regression procedures and  $\Phi^{-1}(p_i)$  is the standard normal inverse cumulative distribution function evaluated at  $p_i$ . Equation (17) is the Blom-based plotting position [*Blom*, 1958] developed by *Hirsch and Stedinger* [1987] for censored data sets. *Helsel and Cohn* [1988] showed that the choice of plotting position is not particularly important when estimating the moments of a normal distribution using censored data sets by (18). The Blom-based plotting position was selected since Blom's plotting position is an approximation to the unbiased plotting position for the normal distribution [*Cunnane*, 1978].

The estimators  $\hat{\mu}_i$  and  $\hat{\sigma}_i$  in (18) are often referred to as log-probability plot regression (LPPR) estimators [*Gilliom and Helsel*, 1986; *Kroll and Stedinger*, 1996] and were originally suggested by *Gupta* [1952]. *Gilliom and Helsel* [1986] compared LPPR estimators to a variety of other estimators including maximum likelihood estimators and found LPPR estimators generally performed well. *Kroll and Stedinger* [1996] also recommended LPPR quantile estimators for low to moderate censoring for data drawn from a lognormal distribution. For each site, once  $\hat{\mu}_i$  and  $\hat{\sigma}_i$  were estimated, an at-site estimate of  $\ln(Q_{7,10})$ ,  $\bar{Y}$ , was calculated as

$$\tilde{Y}_i = \hat{\mu}_i + z_p \hat{\sigma}_i \quad (19)$$

If two or fewer uncensored annual minimum flows were present at a site, the regression analysis indicated by (18) was not performed and the 10-year low flow quantile at the site was assumed to be censored.

### 3.3. Censoring of At-Site Quantile Estimates

One would expect  $Q_{7,10}$  at-site estimates below the measurement threshold to be less precise than above threshold estimates because the former represent an extrapolation below the range of observed data. Results in *Kroll and Stedinger* [1996] illustrate this phenomena. Moreover, censored annual minimum flows may actually be zero, in which case the real  $Q_{7,10}$  may be zero. In this experiment,  $Q_{7,10}$  estimates below 0.1 cfs were censored, which means they were treated as “zeros.”

### 3.4. Regression Model Estimates

Once at-site quantile estimates,  $\tilde{Y}_i$ , have been estimated at all sites, each of the four regression methods were used to estimate the parameters in the model

$$\tilde{Y}_i = \hat{\alpha} + \hat{\beta} \ln(A_i) + \varepsilon_i \quad (20)$$

For the OLSD, TOBIT, and WTOBIT methods, the log-space quantile estimate for site  $i$  with drainage area  $A_i$  was calculated as

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta} \ln(A_i) \quad (21)$$

For the OLSC method, four values of the constant were examined:  $c = 0.01, 0.1, 0.5,$  and  $1.0$  cfs. These represent a range of values less than, equal to, and greater than the censoring threshold of 0.1 cfs. For the OLSC method it is necessary to remove the constant that was added to all the at-site quantiles from the quantile estimates associated with the fitted model. The log-space quantile estimate for site  $i$  with drainage area  $A_i$  for the OLSC method was calculated as

$$\hat{Y}_i = \ln \{ \exp [\hat{\alpha} + \hat{\beta} \ln(A_i)] - c \} \quad (22)$$

Several papers have addressed the bias in log-transformed regression models [*Bradru and Mundlak*, 1970; *Miller*, 1984; *Koch and Smillie*, 1986; *Ferguson*, 1986; *Cohn et al.*, 1989; *McCuen et al.*, 1990]. After performing ordinary least square regression in logarithmic space, a real-space estimator given by  $\hat{Q}_i = \exp(\hat{Y}_i)$  is “median unbiased,” but the expected value of  $\hat{Q}_i$  is a biased estimate of  $Q_i$ . A mean unbiased estimator is given by  $\hat{Q}_i = \exp[\hat{Y}_i + (\sigma_\varepsilon^2/2)]$ , where  $\sigma_\varepsilon^2$  is the true model error variance. In low-flow regional regression analyses, which implement ordinary least squares, the estimate of  $\sigma_\varepsilon^2$  includes both the model error variance and the sampling error variance associated with at-site quantile estimators. One could obtain an estimate of the model error variance using an estimate of the average sampling error associated with the at-site quantile estimators at stations used in the regression and an estimate of the average interstation correlation coefficient [*Hardison*, 1969; *Stedinger and Tasker*, 1986b]. By analyzing the performance of log-space estimators, the problem of transformation bias is avoided for the OLSD, TOBIT, and WTOBIT estimators. To be consistent, the OLSC quantile estimator is transformed back to log space.

### 3.5. Estimation of Sampling Error for WTOBIT

Implementation of the WTOBIT model requires an estimate of the sampling error associated with the at-site quantile estimators,  $\Sigma_{ii}$ . *Stedinger and Tasker* [1985] estimated the sampling error for complete normal samples as

$$\Sigma_{\text{COMPLETE}} = \sigma_i^2 \left( 1 + \frac{z_p^2}{2} \right) / n_i \quad (23)$$

where  $n_i$  is the length of record at site  $i$ ,  $z_p$  is the frequency factor associated with the  $100p$  percentile of a standard normal distribution, and  $\sigma_i^2$  is an estimate of the variance of the log annual minimum flows. Our experiments also require an estimator for censored samples.

Using Monte Carlo simulation, *Kroll and Stedinger* [1996] produced estimates of the sampling variance ( $\Sigma_{\text{LPPR}}$ ) associated with LPPR at-site estimators of  $\ln(Q_{7,10})$  for data drawn from a lognormal distribution. Based on those results, our weighted Tobit model used the approximation

$$\Sigma_{\text{APPROX}} = \Sigma_{\text{COMPLETE}} \exp \left( \frac{kp}{100} \right) \quad (24)$$

where  $p$  is the censoring probability and  $k$  is a constant estimated by minimizing

$$\sum \left( \frac{\Sigma_{\text{APPROX}} - \Sigma_{\text{LPPR}}}{\Sigma_{\text{LPPR}}} \right)^2 \quad (25)$$

where the summation in (25) is over cases with censoring at the 0th, 10th, 20th, and 40th percentiles and coefficients of variation (CV) of 0.25, 0.5, 1.0, and 2.0. An approximation based on (24) was developed for samples of size  $n = 10$ , and another approximation was developed for samples of size  $n = 50$ . For  $n = 10$ ,  $k = 3.2$ , and for  $n = 50$ ,  $k = 2.4$ . In the Monte Carlo experiments,  $p$  was estimated as  $n_c/n_i$ , where  $n_i$  is the record length at site  $i$ , and  $n_c$  is the number of censored annual minimum flows at that site. The percent deviations [ $100(\Sigma_{\text{APPROX}} - \Sigma_{\text{LPPR}})/\Sigma_{\text{LPPR}}$ ] were always less than  $\pm 6\%$  for all censoring percentiles and CVs [see *Kroll*, 1996]. In low-flow regional regression models, the model error variance tends to be larger than the sampling error variance, so the impact of using the approximation in (24) should be minor.

The WTOBIT method requires an estimate of  $\sigma_i$  to obtain an estimate of  $\Sigma_{\text{COMPLETE}}$  in (23). The use of  $\hat{\sigma}_i$  from (18) would produce weights that would be correlated with  $\hat{Y}_i$ . Following the procedures suggested by *Tasker and Stedinger* [1989],  $\sigma_i$  were smoothed by regressing  $\hat{\sigma}_i$  from (18) against  $\ln(A_i)$ . These smoothed  $\sigma_i$  estimates were then used to obtain the weights required by WTOBIT with (23) and (24).

### 3.6. Performance Measures

For each set of parameters, experiments were run until 1000 acceptable data sets were generated. Replicates where no  $Q_{7,10}$  values are censored were discarded from the experiment because our interest is in the performance of the procedures with censored observations. Instances where no  $Q_{7,10}$  values were censored occurred in a maximum of 15 out of 1015 replicates and usually did not occur at all.

The performance of the four regression techniques are described by how well the above-threshold quantile estimates produced by the regression procedures compare to the true values. The performance parameter used to analyze this relationship was the log-space mse. A log-space mse weighs per-

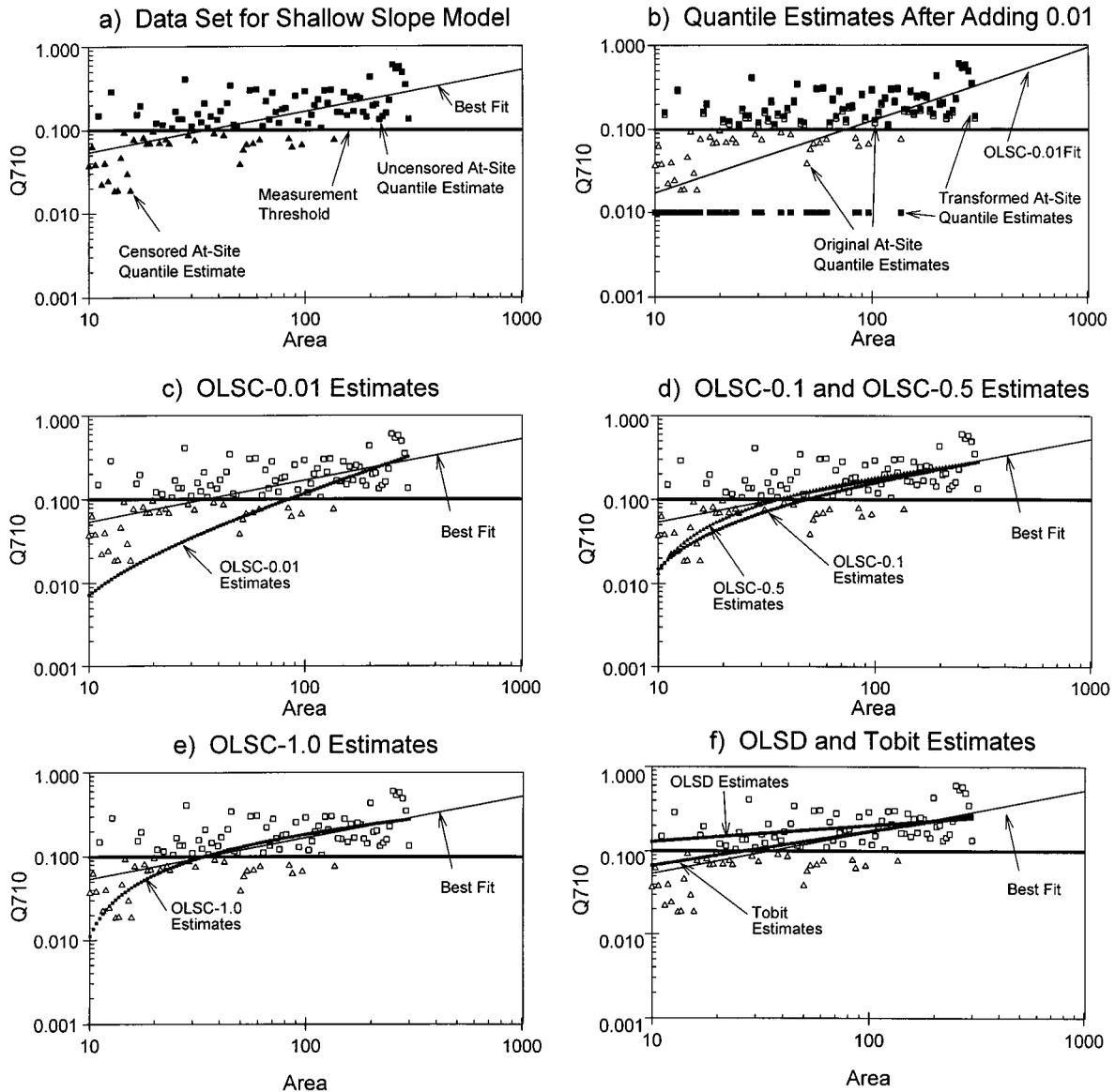


Figure 1. Visual examination of OLSC, OLSD, and TOBIT fits for shallow slope model.

centage over- and under-estimation equally [see *Kroll and Stedinger, 1996*]. In practice one would mainly be interested in improving estimators of the above threshold quantiles because quantile estimators below the censoring threshold are generally treated as zero. For the 25 sites ranked in ascending order by drainage area, where the first  $N_c$  of the sites have  $E[Y_i] = \alpha + \beta A_i < \ln(T)$ , the mse was calculated as the average over 1000 replicates of

$$\text{mse} = \frac{\sum_{i=N_c+1}^{25} (\hat{Y}_i - E[Y_i])^2}{25 - N_c} \quad (26)$$

## 4. Results

In section 4.1 below, data sets are used to visually examine how the OLSC, OLSD, and TOBIT techniques are likely to perform. In section 4.2 the Monte Carlo simulation results

compare the performance of OLSC, OLSD, TOBIT, and WTOBIT estimators.

### 4.1. Visual Examination of OLSC and OLSD Fitting Procedures

The performance of the OLSC and OLSD methods are first examined using data generated from the shallow slope model with  $\alpha_\mu = -2.8$  and a medium model error variance [ $\sigma_\varepsilon^2 = (0.5)^2$ ]. Annual minimum flows were generated at a number of sites in the region. At-site quantiles were then estimated using the generated annual minimum flows, and the regression techniques were employed using the at-site quantile estimates. With 25 sites in a region, using different random data sets, one would expect to observe variations in the at-site quantile estimates about the true values of the quantile. To avoid possible variations in the fitted model, the number of sites in the region was increased to 100 with drainage areas evenly spaced in logarithmic units from 10 to 300 square miles.

Figure 1a is a plot of the 100 at-site quantile estimates versus

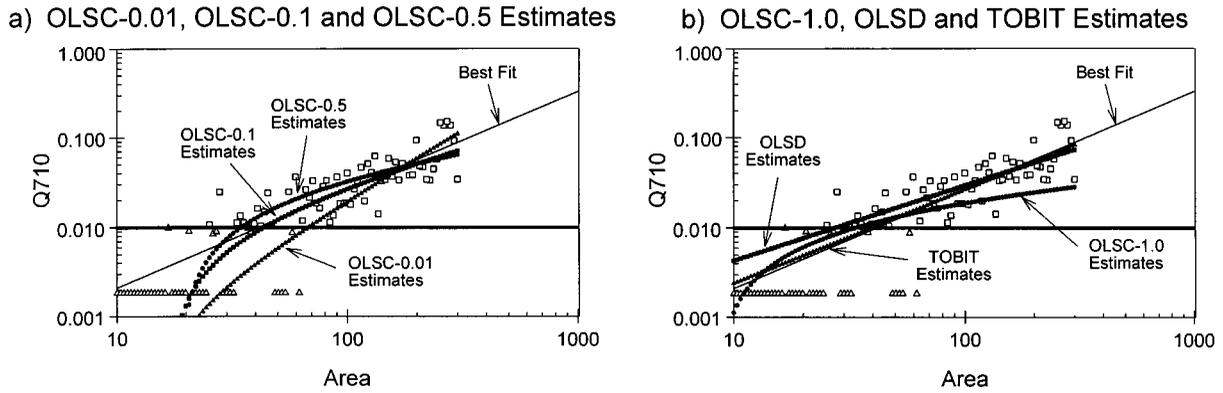


Figure 2. Visual examination of OLS, OLSD, and TOBIT fits for medium slope model.

drainage area for the shallow slope model. The squares represent the uncensored at-site quantile estimates, and the triangles represent the censored at-site quantile estimates which were set to zero. If one could measure all flows regardless of their magnitude, then all annual minimum flows at all sites would be recorded as nonzero and no censoring would occur; in this case, one could obtain uncensored at-site quantile estimates at all sites and fit a regression line to these values. That line is denoted “best fit” in Figure 1a.

For the OLSC-0.01 method, the constant 0.01 is added to all at-site quantile estimates in real space. The dark squares in Figure 1b represent the at-site quantiles in log-space after the constant was added. The OLSC-0.01 method uses OLS regression to fit a line to these transformed at-site quantiles. That line is also plotted in Figure 1b. The light squares and light triangles in Figure 1b represent the original uncensored and censored at-site quantiles from Figure 1a. Quantile estimates for the OLSC-0.01 regression technique are obtained by subtracting 0.01 in real space from the OLSC-0.01 regression line. Those estimates are plotted in Figure 1c. The OLSC-0.01 method does not fit the data very well.

For the OLSC-0.1 method, 0.1 is added to all at-site quantile estimates in real space, and a regression line is fit to the transformed log-space at-site quantile estimates. The resulting quantile estimates for the OLSC-0.1 regression method are found by subtracting 0.1 from the regression line. These estimates are shown in Figure 1d. A similar technique was used for the OLSC-0.5 (Figure 1d) and OLSC-1.0 (Figure 1e) methods. All of the OLSC methods fit the below threshold quantiles poorly. As discussed previously, in practice, one is more concerned with the fit to the above threshold quantiles. For this data set, the OLSC-0.5 and OLSC-1.0 methods appear to fit the above threshold quantile reasonably well.

The OLSD method uses only the uncensored at-site quantile estimates. Figure 1f contains a plot of the OLSD regression fit to the at-site quantile estimates for data generated from the shallow slope model. The performance of the OLSD method is influenced only by at-site quantile estimates which are above the measurement threshold. Above threshold, at-site quantile estimates at sites with small drainage areas are more likely to occur when the slope of the model is shallow and the model error variance is large because the probability of an uncensored at-site quantile estimate is larger in these cases. When such estimates are present, the OLSD method overestimates the intercept term,  $\alpha$ , of the regional model. In Figure 1f the

OLSD method provides a poor description of the above threshold quantiles.

The TOBIT method uses both the censored and uncensored at-site quantile estimates in fitting a regression line without adding a constant to the at-site quantile estimates as the OLS methods does. Figure 1f also contains a plot of the regression line for the TOBIT method for data generated from a shallow slope model. The TOBIT method provides a good fit to both the above and below threshold quantiles.

Figure 2a is a plot of at-site quantile estimates versus drainage area for a 100 site data set using the medium slope model with  $\alpha_\mu = -6.1$  and  $\sigma_\epsilon^2 = (0.5)^2$ . Many of the censored observations are plotted as the same value [ $\exp(-4)$ ]. At these sites, two or fewer annual minimum flows were above the measurement threshold, and thus an at-site quantile estimate was not computed for these sites. The quantile estimates for the OLSC-0.01, OLSC-0.1, and OLSC-0.5 methods are plotted in Figure 2a, with quantile estimates for the OLSC-1.0 method plotted in Figure 2b. In contrast to the shallow slope model where OLSC-0.5 and OLSC-1.0 were the best OLS methods, for the medium slope model, the OLSC-0.1 method appears to fit the above threshold quantiles better than the other OLS methods. The OLSC-0.01 method is again the worst.

Figure 2b also contains the fit of the OLSD and TOBIT methods to the medium slope data. Because no at-site quantiles estimates associated with the smallest drainage areas are above the measurement threshold, the OLSD method fits the above threshold quantiles much better than it did with the shallow slope model. The TOBIT method again provides the best approximation to the best fit model.

#### 4.2. Results of Monte Carlo Simulation

In section 4.1 we examined quantile estimates with two data sets that included 100 sites. In this section we examine how the estimation methods perform in a Monte Carlo simulation of a 25-site region. To examine the relative performance of each of the estimators compared to the best estimator for a particular statistic and a particular set of experimental parameters, the efficiency of each estimator was calculated as

$$\text{Efficiency} = \frac{\text{mse}_{\text{Best Estimator}}}{\text{mse}_{\text{Estimator}}} \quad (27)$$

where the mses were computed using (26). The performance ratio of the estimator with the smallest mse will be equal to 1,

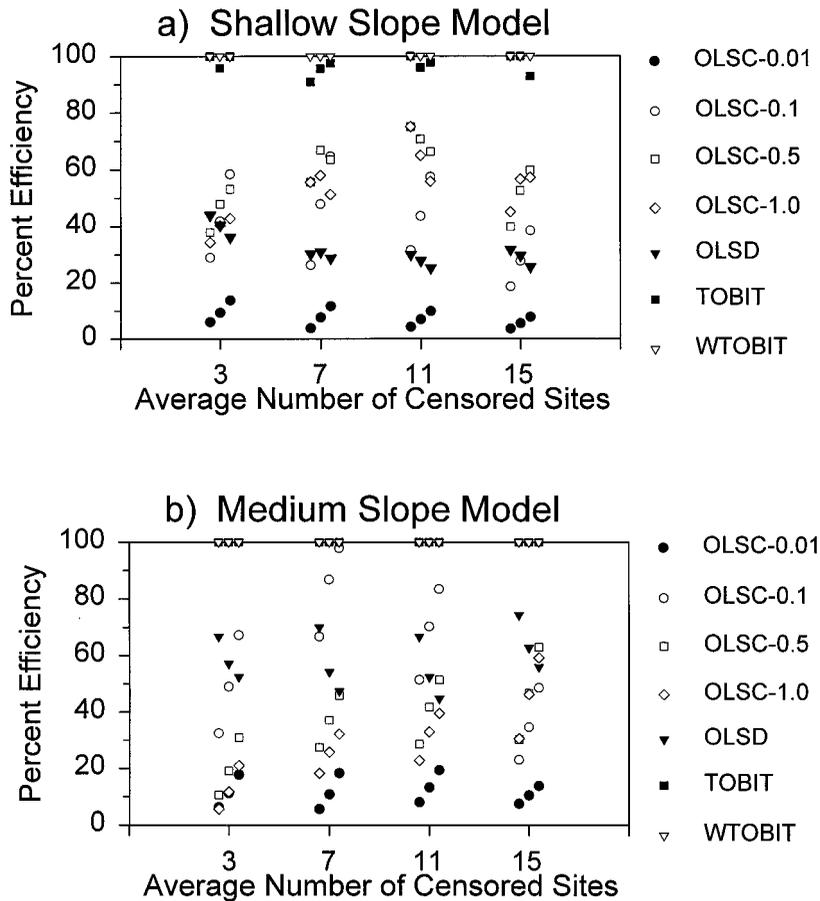


Figure 3. Efficiency of estimators for shallow and medium slope models.

and the performance ratio of the other estimators will be some value between zero and 1. The closer the performance ratio is to 1, the better the estimator is performing relative to the best estimator.

Figure 3a is a plot of the efficiency of the estimators for the shallow slope model. The results are grouped into four sets of three columns. The column on the left corresponds to  $\sigma_\varepsilon = 0.3$ , the middle column to  $\sigma_\varepsilon = 0.5$ , and the right column to  $\sigma_\varepsilon = 0.7$ . As expected, the TOBIT and WTOBIT methods clearly dominate the other techniques. The WTOBIT method sometimes performs slightly better than the TOBIT method, though this difference is modest.

When viewing the fits of the OLSC and OLSD methods to the data set from the shallow slope model data in section 4.1, we saw that the OLSC-0.5 and OLSC-1.0 methods fit the above threshold quantiles better than the other methods. This result is also reflected in Figure 3a. The OLSC-0.01 method is the worst-fitting technique having an efficiency less than 20%.

Figure 3b is a plot of the efficiency of the estimators for the medium slope model. Again the Tobit methods dominate the other techniques. The WTOBIT method produced no increase in mse efficiency over the TOBIT method for the medium slope model results. In many cases the OLSD method performs better than all the OLSC methods. This is because in the medium slope model it is relatively unlikely that at-site quantiles associated with small drainage area sites will be above the measurement threshold when the median value of the quantile is below the censoring threshold.

## 5. Conclusions

In this experiment the performance of estimation techniques for regional regression analyses with censored data were examined. One technique, OLSC, added a constant to all at-site quantile estimates in real space before performing a log linear regression analysis. This method was examined for constants less than, equal to, and larger than the measurement threshold. A second method, OLSD, discards all censored at-site quantile estimates and performs a log linear regression analysis using the remaining uncensored at-site quantile estimates. The third method, TOBIT, uses a maximum likelihood technique to estimate the model parameters. In the likelihood function the censored data are represented as a probability of being less than the censoring threshold. The fourth method, WTOBIT, is a weighted Tobit method which accounts for the heteroscedastic nature of the residuals due to differences in the sampling error variance associated with at-site quantile estimators. These estimation methods are compared using both Monte Carlo simulation and visual analysis.

The performance criterion used in the Monte Carlo experiment to compare these methods was the log-space mean square error (mse) of model quantile estimators. In practice, one would mainly be interested in improving estimators of quantiles which are greater than the measurement threshold because quantile estimators below the censoring threshold are generally treated as zero. In this experiment the estimation techniques are compared for a region with a shallow slope and

with a medium slope. In multivariate models one would potentially have some parameters that would be representative of shallow slopes while others were medium or even high sloped parameters.

In general, for the estimators considered in this experiment the following occur:

1. Regardless of the amount of censoring or the slope of the underlying regional model, the Tobit methods always produced estimators with substantially smaller mses than the OLSD and OLSC quantile estimators.

2. The weighted Tobit method, WTOBIT, had little advantage over the TOBIT method. This result was because the sampling error variances employed with the WTOBIT method were small compared to the model error variance, and thus the weights were nearly constant.

3. The performance of the OLSD method depended on the occurrence of at-site quantile estimates above the measurement threshold at small drainage area sites whose median quantile value was below the censoring threshold. When these above threshold estimates occurred, the OLSD method overestimated the intercept parameter in the regression equation. This was more likely with the shallow slope model, especially for regions with a large model error variance.

4. For the shallow slope model, the OLSC-0.5 and OLSC-1.0 methods tended to produce quantile estimators with a smaller average mse than the OLSD and OLSC-0.1 methods. For the medium slope model, the OLSD and OLSC-0.1 methods tended to produce quantile estimators with a smaller average mse than the OLSC-0.5 and OLSC-1.0 methods. The OLSC-0.01 method performed very poorly in all cases.

Based on these results, we recommend the following:

1. Ideally with censored at-site quantiles, it is best to use a Tobit model to estimate nonzero quantiles, though this procedure is more computationally intensive than the OLSD and OLSC methods. The Tobit method is the only technique which performed well in all cases examined. Computer packages such as SAS [SAS Institute, 1989] and LIMDEP [Greene, 1995] are available to estimate the parameters of a Tobit model.

2. While the OLSC method may perform satisfactorily if the correct constant is employed with this method, it is difficult to know which constant is best, and thus this method is not reliable.

3. For a small amount of censoring, the OLSD method can be satisfactory for cases with a small model error variance.

4. Use of the WTOBIT method is generally not warranted with low-flow data sets because it is more computationally intensive than the TOBIT method and little or no reduction in the mse of the quantile estimators is achieved.

**Acknowledgments.** The authors would like to thank Glenn Moglen and an anonymous reviewer whose comments helped clarify a number of important issues. This material is based upon work partially supported by Cooperative State Research Service, USDA, under project 97CRMS06102 and the U.S. Department of the Interior through the New York State Water Resources Institute, under project 1434-HQ-96-GR02688.

## References

Amemiya, T., Regression analysis when the dependent variable is truncated normal, *Econometrica*, 41, 997–1016, 1973.  
 Amemiya, T., *Advanced Econometrics*, Harvard Univ. Press, Cambridge, Mass., 1985.  
 Arihood, L. D., and D. R. Glatfelter, Method for estimating low-flow

characteristics of ungauged streams in Indiana, *U.S. Geol. Surv. Water Supply Pap.*, 2372, 1991.  
 Bingham, R. H., Regionalization of low-flow characteristics of Tennessee streams, *U.S. Geol. Surv. Water Resour. Invest. Rep.*, 85-4191, 1986.  
 Blom, G., *Statistical Estimates and Transformed Beta Variables*, John Wiley, New York, 1958.  
 Bradu, D., and Y. Mundlak, Estimation in lognormal linear models, *J. Am. Stat. Assoc.*, 65, 198–211, 1970.  
 Chapra, S. C., and R. P. Canale, *Numerical Methods for Engineers*, McGraw-Hill, New York, 1985.  
 Cohn, T. A., L. L. DeLong, E. J. Gilroy, R. M. Hirsch, and D. K. Wells, Estimating constituent loads, *Water Resour. Res.*, 25, 937–942, 1989.  
 Cunnane, C., Unbiased plotting positions: A review, *J. Hydrol.*, 37, 205–222, 1978.  
 Dingman, S. L., and S. C. Lawlor, Estimating low-flow quantile from drainage-basin characteristics in New Hampshire and Vermont, *Water Resour. Bull.*, 31, 243–256, 1995.  
 Durrans, S. R., and S. Tomic, Regionalization of low-flow frequency estimates: An Alabama case study, *Water Resour. Bull.*, 32, 23–37, 1996.  
 Ferguson, R. I., River loads underestimated by rating curves, *Water Resour. Res.*, 22, 74–76, 1986.  
 Gilliom, R. J., and D. R. Helsel, Estimation of distributional parameters for censored trace level water quality data, 1, Estimation techniques, *Water Resour. Res.*, 22, 135–146, 1986.  
 Greene, W. H., *LIMDEP Version 7.0*, Econometric Software, Inc., Plainview, N. Y., 1995.  
 Gupta, A. K., Estimation of the mean and standard deviation of a normal population from a censored sample, *Biometrika*, 39, 260–273, 1952.  
 Hammett, K. M., Low-flow frequency analysis for streams in west-central Florida, *U.S. Geol. Surv. Water Resour. Invest. Rep.*, 84-4299, 1984.  
 Hardison, C. H., Accuracy of streamflow characteristics in geological survey research, *U.S. Geol. Surv. Prof. Pap.*, 650-D, 1969.  
 Helsel, D. R., and T. A. Cohn, Estimation of descriptive statistics for multiply censored water quality data, *Water Resour. Res.*, 24, 1997–2004, 1988.  
 Hirsh, R. M., An evaluation of some record reconstruction techniques, *Water Resour. Res.*, 15, 1781–1790, 1979.  
 Hirsh, R. M., A comparison of four stream flow record extension techniques, *Water Resour. Res.*, 18, 1081–1088, 1982.  
 Hirsh, R. M., and J. R. Stedinger, Plotting positions for historical floods and their precision, *Water Resour. Res.*, 23, 715–727, 1987.  
 Jennings, M. E., W. O. Thomas Jr., and H. C. Riggs, Nationwide summary of geologic survey regional regression equations for estimating magnitude and frequency of floods for ungauged sites, 1993, *U.S. Geol. Surv. Water Resour. Invest. Rep.*, 94-4002, 1994.  
 Johnston, J., *Econometric Methods*, McGraw-Hill, New York, 1972.  
 Judge, G. G., W. E. Griffiths, R. C. Hill, H. Lutkepohl, and T. C. Lee, *The Theory and Practice of Econometrics*, chap. 18, John Wiley, New York, 1985.  
 Koch, R. W., and G. M. Smillie, Bias in hydrologic prediction using log-transformed regression models, *Water Resour. Bull.*, 22, 717–723, 1986.  
 Kroll, C. N., Censored data analyses in water resources, Ph.D. thesis, Cornell Univ., Ithaca, N. Y., 1996.  
 Kroll, C. N., and J. R. Stedinger, Low-flow frequency analysis with censored data, paper presented at 21st Annual American Society of Civil Engineers Conference and Symposium: Water Resources Planning and Management, Denver, Colorado, 1994.  
 Kroll, C. N., and J. R. Stedinger, Estimation of distributional moments and quantiles for censored data, *Water Resour. Res.*, 32, 1005–1012, 1996.  
 Kroll, C. N., and J. R. Stedinger, Regional hydrologic analysis: Ordinary and generalized least squares revisited, *Water Resour. Res.*, 34, 121–128, 1998.  
 Lawal, S. A., and W. E. Watt, Non-zero lower limit in low flow frequency analysis, *Water Resour. Bull.*, 32, 1159–1166, 1996.  
 Liu, S., J.-C. Lu, D. W. Kolpin, and W. Q. Meeker, Analysis of environmental data with censored observations, *Environ. Sci. Technol.*, 31, 3358–3362, 1997.  
 Lu, J. C., S. Liu, M. Yin, and J. M. Hughes-Oliver, Modeling restricted bivariate censored lowflow data, *Econometrica*, in press, 1998.  
 Ludwig, A. H., and G. D. Tasker, Regionalization of low-flow charac-

- teristics of Arkansas streams, *U.S. Geol. Surv. Water Resour. Invest. Rep.*, 93-4013, 1993.
- Lui, S., and J. R. Stedinger, Low stream flow frequency analysis with Ordinary and Tobit Regression, paper presented at 18th Annual American Society of Civil Engineers Conference and Symposium: Water Resources Planning and Management and Urban Water Resources, New Orleans, La., 1991.
- McCuen, R. H., R. B. Leahy, and P. A. Johnson, Problems with logarithmic transformations in regression, *J. Hydraul. Eng.*, 116, 414-428, 1990.
- Miller, D. M., Reducing transformation bias in curve fitting, *Am. Stat.*, 38, 124-126, 1984.
- Olsen, R. J., Note on the uniqueness of the maximum likelihood estimator for the Tobit model, *Environmetrics*, 46, 1211-1215, 1978.
- Parker, G. W., Methods for determining selected flow characteristics for streams in Maine, *U.S. Geol. Surv. Open File Rep.*, 78-871, 1977.
- Riggs, H. C., Frequency curves, U.S. Geol. Surv. Surface Water Techniques Series, Book 4, Chapter A2, U.S. Geol. Surv., Reston, Va., 1968.
- Riggs, H. C., Low-Flow Investigations, U.S. Geol. Surv. Techniques of Water Resource Investigations, Book 4, Chapter B1, U.S. Geol. Surv., Reston, Va., 1972.
- Riggs, H. C., et al., Characteristics of low flows, *J. Hydraul. Eng.*, 106, 717-731, 1980.
- SAS Institute, *SAS/STAT User's Guide Version 6*, 4th ed., pp. 997-1024, SAS Institute Inc., Cary, N. C., 1989.
- Stedinger, J. R., and G. D. Tasker, Regional hydrologic analysis, 1, Ordinary, weighted, and generalized least squares compared, *Water Resour. Res.*, 21, 1421-1432, 1985.
- Stedinger, J. R., and G. D. Tasker, Regional hydrologic analysis, 1, Ordinary, weighted, and generalized least squares compared, *Water Resour. Res.*, 22, 844, 1986a. (Correction, *J. Geophys. Res.*, 21, 1421-1432, 1985.)
- Stedinger, J. R., and G. D. Tasker, Regional hydrologic analysis, 2, Model-error estimators, estimation of sigma and log-Pearson type 3 distribution, *Water Resour. Res.*, 22, 1487-1499, 1986b.
- Stedinger, J. R., and W. O. Thomas Jr., Low-flow frequency estimation using base-flow measurements, *U.S. Geol. Surv. Open File Rep.*, 85-95, 1985.
- Tasker, G. D., Hydrologic regression with weighted least squares, *Water Resour. Res.*, 16, 1107-1113, 1980.
- Tasker, G. D., A comparison of methods for estimating low flow characteristics of streams, *Water Resour. Bull.*, 23, 1077-1083, 1987.
- Tasker, G. D., and J. R. Stedinger, An operational GLS model for hydrologic regression, *J. Hydrol.*, 111, 361-375, 1989.
- Thomas, D. M., and M. A. Benson, Generalization of streamflow characteristics from drainage-basin characteristics, *U.S. Geol. Surv. Water Supply Pap.*, 1975, 1970.
- Thomas, M. P., and M. A. Cervione, A proposed streamflow data program for Connecticut, *Conn. Water Resour. Bull.*, 23, 1-18, 1970.
- Tobin, J., Estimation of relationships for limited dependent variables, *Econometrica*, 26, 24-36, 1958.
- Vogel, R. M., and C. N. Kroll, Low-flow frequency analysis using probability-plot correlation coefficients, *J. Water Resour. Plann. Manage.*, 115, 338-357, 1989.
- Vogel, R. M., and C. N. Kroll, Generalized low-flow frequency relationships for ungauged sites in Massachusetts, *Water Resour. Bull.*, 26, 241-253, 1990.
- Vogel, R. M., and C. N. Kroll, Regional geohydrologic-geomorphic relationships for the estimation of low-flow statistics, *Water Resour. Res.*, 28, 2451-2458, 1992.
- Vogel, R. M., and J. R. Stedinger, Minimum variance streamflow record augmentation procedures, *Water Resour. Res.*, 21, 715-723, 1985.

---

C. N. Kroll, Environmental Resources and Forest Engineering, College of Environmental Science and Forestry, State University of New York, 312 Bray Hall, 1 Forestry Drive, Syracuse, NY 13210. (e-mail: cnkroll@mailbox.syr.edu)

J. R. Stedinger, School of Civil and Environmental Engineering, Cornell University, Ithaca, NY 14850.

(Received December 1, 1997; revised August 20, 1998; accepted August 21, 1998.)